

# **Coursera Capstone: The battle of Neighborhoods (Schools In Manchester)**

## **Introduction: Business Problem:**

This project identifies and recommends the nearest schools and school categories in neighborhoods. People especially families may move from one settlement to the other without considering the status of schools and their categories within the city and its neighborhoods.

The project targets families and students who will need to consider schools and their categories within the neighborhoods of a city before making their choices of settlement.

With regards to development and building of school structures in a city, the government will need to analyze schools and their categories within the neighborhoods to make the best choices for residents.

Private sectors, investors and business institutions are also beneficiaries. These entities can secure the right locations to invest by making the right decision on which categories of schools will be of a high demand in various neighborhoods.

## **WHO WILL BE INTERESTED?**

As I have explained earlier, the target audience will be the government agencies looking to build school infrastructure, the private sector agencies looking to build schools in areas with good potential and the people looking for houses in different neighborhoods to identify which ones will be closer to nearby schools.

## **Data Requirement:**

Based on the problem definition above, factors to be considered are:

- Number of schools and their categories in a neighborhood
- Radius distance of neighborhoods nearest to schools

The following data sources will be required to extract relevant data:

- A list of neighborhoods of Manchester from Wikipedia
- A geocoder library for extracting latitude and longitude coordinates
- FourSquare API for searching for schools within these neighborhoods and providing their various latitudes and longitude coordinates

### **Preprocessing the Data:**

The data had some irrelevant records in the dataset which I removed. For instance, I searched for schools using the FourSquare API and the results included school cafeteria and entertainment parks. I also eliminated duplicates which came up as a result of schools present in neighborhoods within a 1000 meter radius.

Below is an image of my pre processed data(showing first five) for analysis:

	PostalCode	Town	Neighborhood	Latitude	Longitude
0	M1	Manchester	Piccadilly, City Centre, Market Street	53.479396	-2.231743
1	M2	Manchester	Deansgate, City Centre	53.474096	-2.251063
2	M4	Manchester	Ancoats, Northern Quarter, Strangeways	53.484998	-2.227603
3	M8	Manchester	Crumpsall, Cheetham Hill	53.520912	-2.241695
4	M9	Manchester	Harpurhey, Blackley	53.511907	-2.208449

### **Methodology:**

First, the data has been retrieved and cleaned as explained previously. The data consists of schools within a radius of 1000m for each neighborhood with a limit set to 100.

Next, we analyze the data by exploring the various categories to which the schools belong.

After this, we explore the distribution in each neighborhood in a more explanatory way using maps for visualization. This gives us a good idea of the distribution of schools.

Then for those neighborhoods without schools, we found the distance to the closest school by incrementing the radius by 100 meters in each iteration until a school is found using Folium.

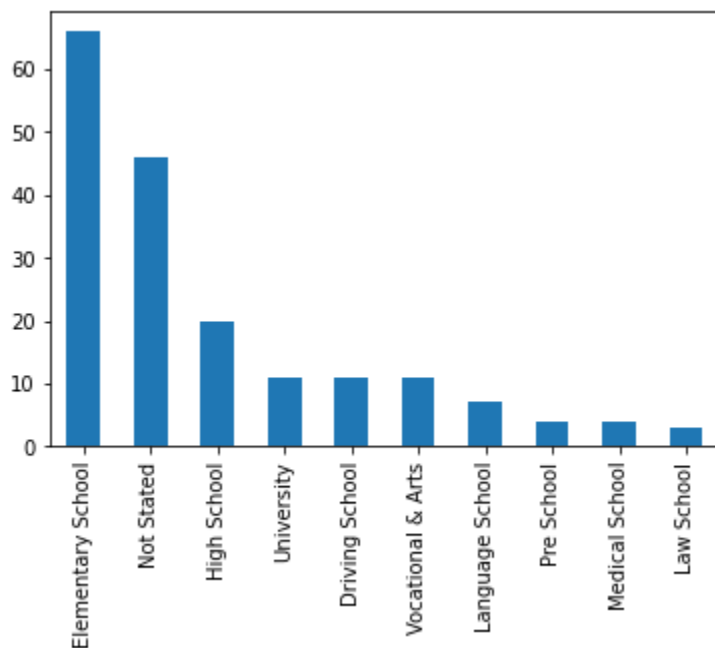
On obtaining this refined data, consisting of the distance to the nearest school for each neighborhood and the number of schools in each neighborhood we cluster this data using **K-Means**. This helps us identify those regions where there is a need for improvement and those neighborhoods which have good facilities in terms of schools.

Finally, we will discuss and conclude which neighborhoods require an improvement of facilities in terms of schools and which regions have very good facilities.

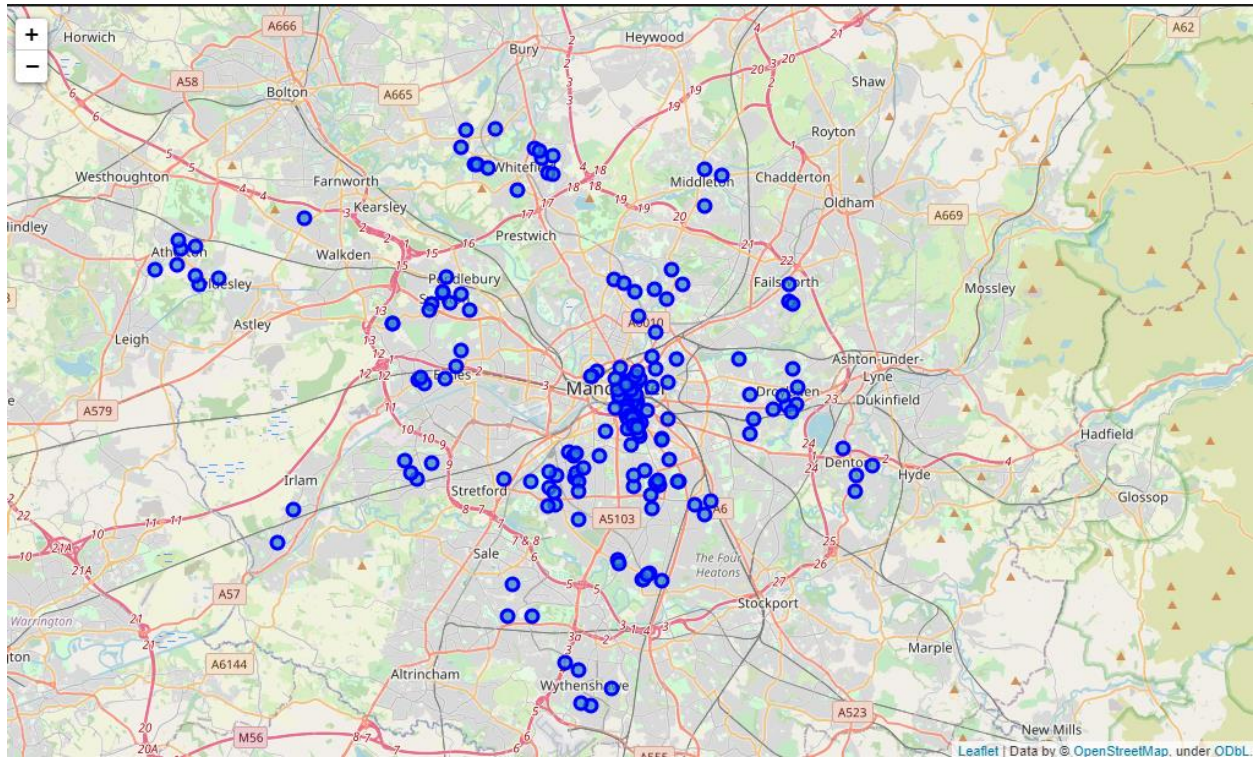
### **Analysis:**

Based on the data obtained from Folium about schools, I explored the categories to which schools belong.

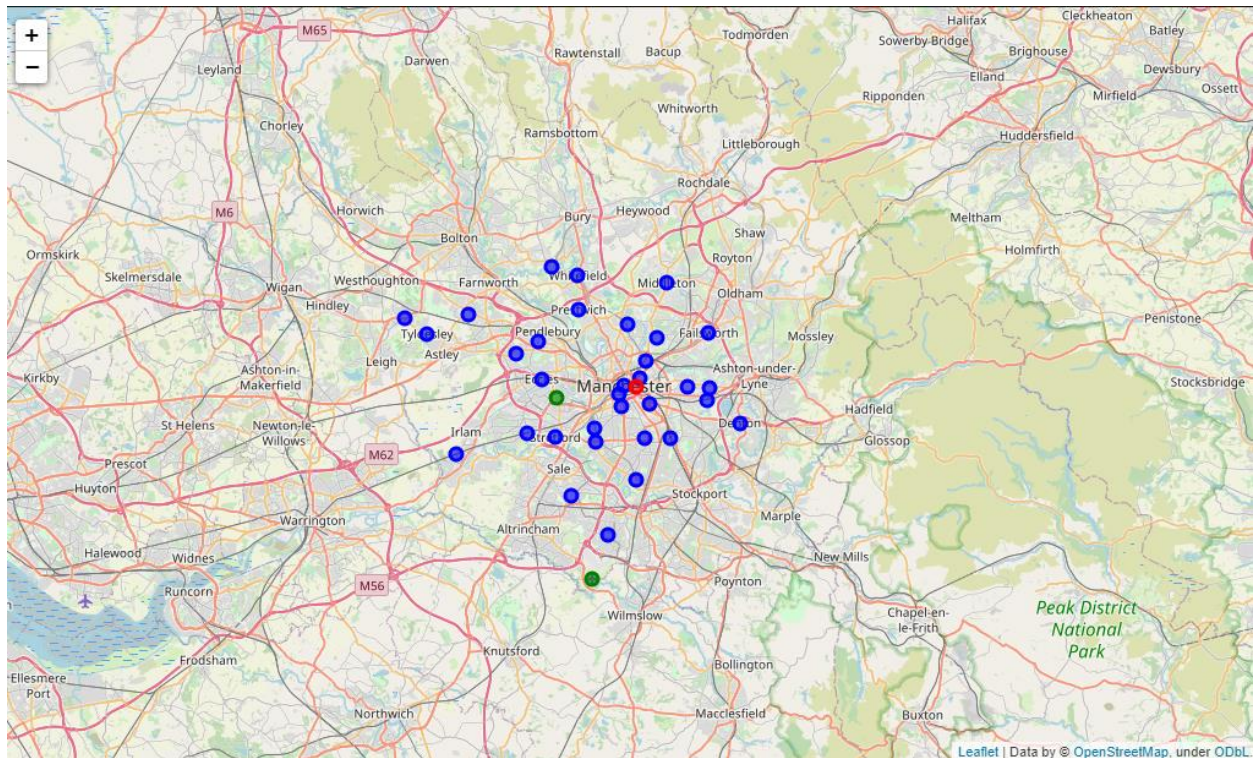
### **Categories:**



## Distribution of Neighborhoods on map:



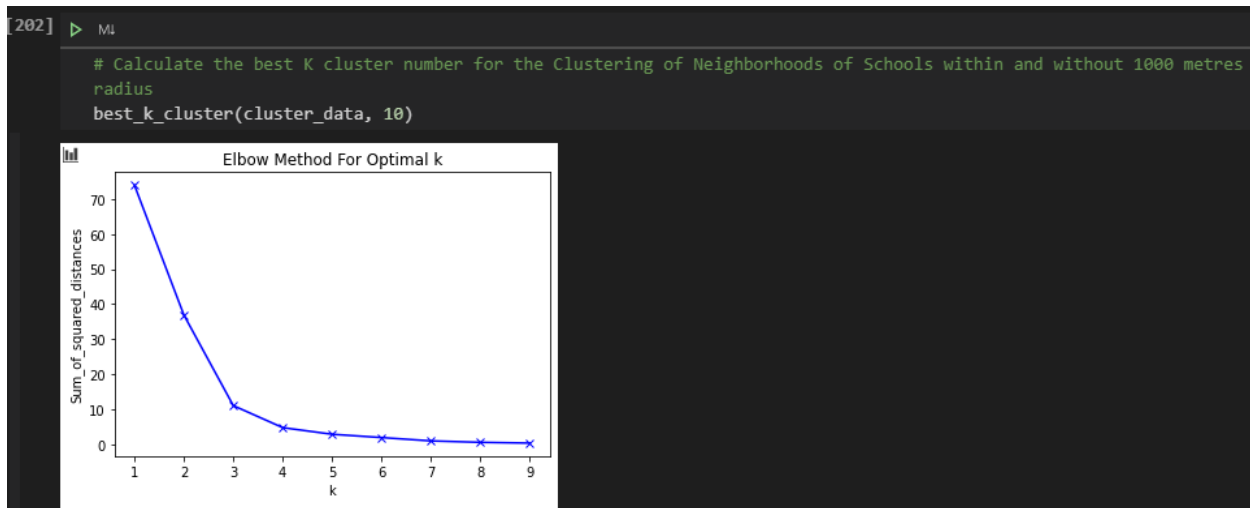
## Distribution of Schools in Neighborhoods on map:





Using **Folium to generate a map**, we can get a good idea of the distribution of schools.

We find the best cluster number to use by finding the mean of the radius of the data within the cluster number range of 1 to 10.



### Clustering Neighborhoods:

We will now cluster the neighborhoods using K-Means clustering algorithm into 3 clusters based on the distance to nearest school and number of schools in each neighborhood.

These features form a good basis for identifying the neighborhoods requiring improvement.

The nearest distance to school is varied between a radius of 1000 meters to 2000 meters. The image below shows the neighborhoods with the nearest schools between 1000 meters and 2000 meters.

```
hoods_without_schools_df['Sch_in_1000m']=0
hoods_without_schools_df.head()
```

	Neighborhood	Latitude	Longitude	Sch_Radius	Sch_in_1000m
0	Trafford Park, The Trafford Centre	53.471752	-2.320928	1600	0
1	Prestwich, Sedgley Park, Simister	53.530427	-2.296019	1100	0
2	Carrington, Partington	47.449720	-99.126224	1100	0
3	Manchester Airport	53.350342	-2.280369	1800	0

After merging with neighborhoods with schools within the 1000 meters radius. The result was:

	Neighborhood	Latitude	Longitude	Sch_Radius	Sch_in_1000m	label
0	Trafford Park, The Trafford Centre	53.471752	-2.320928	1600	0	1
1	Prestwich, Sedgley Park, Simister	53.530427	-2.296019	1100	0	0
2	Carrington, Partington	47.449720	-99.126224	1100	0	0
3	Manchester Airport	53.350342	-2.280369	1800	0	1
4	Abbey Hey, Gorton	53.469858	-2.152198	1000	6	0

Next, we look into the different clusters. The averages of the features for each cluster is as follows.

```
Merged_Hood_Radii_df.groupby('label').mean()
```

	Latitude	Longitude	Sch_Radius	Sch_in_1000m
label				
0	53.303404	-5.121835	1005.882353	4.382353
1	53.411047	-2.300648	1700.000000	0.000000
2	53.479396	-2.231743	1000.000000	34.000000

Obviously, the average of Latitudes and Longitudes has no meaning in this context.

As it is clear, the neighborhoods labeled 0 not only have schools **within 1000m** (which let me remind you is the starting point for nearest distance and it can't get any better) but also have a number of options in terms of schools with an average of **4.38**. These neighborhoods have hence access to excellent education facilities in terms of school level.

The neighborhoods labeled 1 are the ones **requiring improvement** as the average distance to nearest schools is close to **2km** with no schools within 1km of the neighborhood. These are present mostly in the outskirts of the city.

The neighborhoods labeled 2 also have easy access to good education facilities at school level within **1000m radius** with an average of about **34 schools** per neighborhood. Most of the neighborhoods fall under this category.

### **Results and Discussions:**

We have observed the various categories the schools belong to in Manchester, with most of them being more of general schools which provide education from Pre School to University.

Also, we observe that schools providing education only at university level are comparatively less which is expected.

Next, we have observed the distribution of schools and find that mostly they are concentrated towards the center of the city and are less dense in the outskirts. This is clear from the map above.

When someone is looking to find a neighborhood with easy access to good facilities, they should find such neighborhoods closer to the center of the city.

Finally, we have clustered the neighborhoods into 3 clusters and have observed that most of the neighborhoods fall under category with label 2, i.e, the average distance to nearest school for these neighborhoods is 1000 meters and there are about 34 schools on average in these neighborhoods.

These regions consist of potential neighborhoods for the private sector to build schools in, because they don't have many schools but at the same time there will

be many people living in these neighborhoods as they have access to good facilities.

We have also identified the neighborhoods requiring improvement which is the motive of this analysis.

These neighborhoods have a school with an average radius to closest school close to 2000m which was the upper limit set on the distance to nearest school feature.

These are the regions requiring attention in terms of improvement at the school level. The government should concentrate on these areas.

The full list of neighborhoods in each cluster can be found in the Jupyter notebook link I have attached below.

[the battle of neighborhoods](#)

### **Conclusion:**

In this analysis, the distribution of schools in Manchester has been analyzed. Also, the various neighborhoods have been clustered based on their access to school facilities per distance. The neighborhoods requiring improvement have been identified along with those neighborhoods which have the potential for profitable schools. Also, the various categories of schools have been observed.